PUMP Package: Technical Appendix

Kristen Hunter, Luke Miratrix, Kristin Porter

May 29, 2023

Contents

1   Introduction

Our package allows for power calculations across a range of common scenarios that a user might select. Each scenario is categorized by two choices. First, the user chooses the planned experimental design (e.g., clustered data, with randomization within cluster). Second, the user makes a choice of the planned analytic model they would use for the data (e.g. a multilevel model with random impacts). To provide a concrete example, throughout we assume an education setting where we have students (level 1) nested within schools (level 2) nested, in the three-level case, within districts (level 3).

The *design* is characterized by the number of levels of nesting (e.g., students in schools, no districts, would be two levels) and the level of randomization (e.g., randomization of schools would be randomization at level two).

The *model* choices are a bit more complex, and we discuss them in detail below. In particular, for each model we support, we create a taxonomy by noting what modeling choice is used at each level of the model, and how covariates are used.

For all designs and models, we only support estimating the causal effect of a binary treatment.

The outline of this appendix is as follows. In the introduction, we provide notation, a taxonomy for models, an explanation of user-set parameters, and an outline the power estimation strategy. The bulk of the appendix then provides detailed information about each supported scenario, including the assumed model and consequent standard error formula. We then describe the data generating

Param	Description
M	Number of outcomes
$\bar{n}$	Number of level 1 units in each level 2 group (assumed constant across level 2 groups)
J	Number of level 2 units in each level 3 group (assumed constant across level 3 groups)
K	Number of level 3 units
$\bar{T}$	Proportion of the sample that is assigned to the treatment group (assumed constant across groups)
N	Total number of units: $N = \sum_{k=1}^{K} \sum_{j=1}^{J} \bar{n}$
S_{id}	Categorical variable indicating the membership of individual i to a level 2 group
D_{id}	Categorical variable indicating the membership of individual i to a level 3 group
Y_{ijmk}	Observed outcome for unit i in level 2 group j in level 3 group k for outcome m
T_{ijk}	Binary treatment assignment for unit i in level 2 group j in level 3 group k
$\mathbf{V}_{km}$	Vector of level 3 covariates
$g_{3,m}$	Number of level 3 covariates for outcome m
$\mathbf{X}_{jkm}$	Vector of level 2 covariates
$g_{2,m}$	Number of level 2 covariates for outcome m
$\mathbf{C}_{ijkm}$	Vector of level 1 covariates
$g_{1,m}$	Number of level 1 covariates for outcome m

Table 1: Observed quantities.

process that we used for the package validation; this section might also be useful for readers who wish to understand the models and assumptions more deeply. The final two sections provide explicit formula linking the user-specified parameters used for power analysis to the full set of parameters used for data generation, followed by derivations of these formula.

1.1   Notation

To define our models, we first assume a set of observed quantities, shown in Table 1, such as sample sizes, the outcomes, and covariates. For all notation, we use i to index level 1 (individuals), j to index level 2 (schools), k to index level 3 (districts), and m to index outcomes.

We define an impact as the difference in outcome when $T_{ijk} = 1$ compared to when $T_{ijk} = 0$. Define $Y_{ijkm}(T_{ijk} = 1)$, or $Y_{ijkm}(1)$ for short, as the potential outcome given $T_{ijk} = 1$, and $Y_{ijkm}(0)$ as the potential outcome given $T_{ijk} = 0$. We can then define the average treatment impact for outcome m

at different levels:

$$\text{Level 1:} \quad \theta_{1,jkm} = \frac{1}{\bar{n}} \sum_{i=1}^{\bar{n}} \left[Y_{ijkm}(1) - Y_{ijkm}(0) \right] \quad \text{for a given school } j \text{ in district } k. \tag{1}$$

$$\text{Level 2:} \quad \psi_{1,km} = \frac{1}{J} \sum_{j=1}^{J} \theta_{1,jkm} \quad \text{for district } k. \tag{2}$$

$$\text{Level 3:} \quad \Xi_{1,m} = \frac{1}{K} \sum_{k=1}^{K} \psi_{1,km}. \tag{3}$$

For three-level designs, our estimand is the average district treatment impact $\Xi_{1,m}$; for two-level designs the average school impact $\psi_{1,m}$; and for one-level designs the average student impact $\theta_{1,m}$. We only consider person-weighted estimands, where each level 1 unit is equally weighted. If we have different sample sizes within a level, such as having different numbers of students in each school, we could alternatively consider estimands that weight by the size of the school.

We also use group mean outcomes under no treatment at different levels:

$$\text{Level 1:} \quad \theta_{0,jkm} = \frac{1}{\bar{n}} \sum_{i=1}^{\bar{n}} Y_{ijkm}(0) \quad \text{for a given school } j \text{ in district } k. \tag{4}$$

$$\text{Level 2:} \quad \psi_{0,km} = \frac{1}{J} \sum_{j=1}^{J} \theta_{0,jkm} \quad \text{for district } k. \tag{5}$$

$$\text{Level 3:} \quad \Xi_{0,m} = \frac{1}{K} \sum_{k=1}^{K} \psi_{0,km}. \tag{6}$$

We call the means of levels two and three "grand" means because they are the means of means. These quantities are listed in Table 2.

1.2   Model taxonomy

All of our models are variants of regression models. Depending on specification we might have fixed or random effects at different levels. Table 3 lists our notation we use for the latent parameters associated with these models. The residual intercepts and impacts are deviations from the respective group means. For example, $u_{1,jkm}$ would be how School j in District k's average treatment effect for outcome m was different from the average impact in District k, i.e., $u_{1,jkm} = \theta_{1,jkm} - \psi_{1,km}$.

Using our full set of notation, we can create a model taxonomy based on the modeling choices used for all the intercepts and impacts. In particular we determine for each level:

- Whether the level 2 and level 3 intercepts are:
  - fixed ($u_{0,jkm}$ and/or $w_{0,jkm}$ are fixed effects)
  - random ($u_{0,jkm}$ and/or $w_{0,jkm}$ are considered to be Normally distributed, allowing for partial pooling).
- Whether the level 2 and level 3 treatment effects are:

4

Param	Description
$Y_{ijmk}(T_{ijk} = 0)$	Potential outcome for unit i in level 2 group j in level 3 group k for outcome m given control treatment
$Y_{ijmk}(T_{ijk} = 1)$	Potential outcome for unit i in level 2 group j in level 3 group k for outcome m given active treatment
$\Xi_{0,m}$	Grand mean outcome under no treatment across level 3 units for outcome m
$\Xi_{1,m}$	Grand mean impact across level 3 units for outcome m
$\psi_{0,km}$	Grand mean outcome under no treatment across level 2 units in level 3 unit k for outcome m
$\psi_{1,km}$	Grand mean impact across level 2 units in level 3 unit k for outcome m
$\theta_{0,jkm}$	Mean outcome under no treatment for individuals in level 2 unit j in level 3 unit k for outcome m
$\theta_{1,jkm}$	Mean impact on individuals in level 2 unit j in level 3 unit k for outcome m

Table 2: Means and average impacts.

- constant, e.g., all units are modeled as having a single average impact ($u_{1,jkm} = 0$ and/or $w_{1,km} = 0$).

- fixed, with each unit has an individual estimated impact making the $u_{1,jkm}$ or $w_{1,jkm}$ fixed effects constrained to have mean 0, with an additional overall mean impact.

- random, with $u_{1,jkm}$ and $w_{1,km}$ modeled as Normally distributed around an overall mean impact.

On Table 3, the variance terms can be thought of either as the variance of some random effect or as the variance of some fixed effects across the groups in the sample.

In addition, for each level the user can plan to adjust for baseline covariates, unless there are fixed effects at that level or below (if you have a dummy variable for each group in a level, a covariate at that level would be co-linear with those dummies). The coefficient vectors listed on Table 3 would be estimated in models that use the baseline covariates as controls in the estimation process.

1.3   PowerUp!

For users familiar with PowerUp! (Dong and Maynard, 2013; Bulus et al., 2022), Table 4 provides a reference for translating notation between PowerUp! and this document. Note that in some models, PowerUp! includes treatment by covariate interactions, allowing for, in principle, heterogeneous treatment effects correlated with said covariates. We do not allow for this, as including treatment by covariate interactions adds complexity with estimation of average treatment effects, and is unlikely to help with the precision of an overall average impact estimate. We view covariate by treatment interactions as primarily for modeling treatment effect heterogeneity, which is not the goal of this package or project. When this difference between PowerUp! and our approach occurs, it is noted.

Param	Description
$w_{0,km}$	Level 3/District residual intercepts
$w_{1,km}$	Level 3/District residual impacts
$\eta_{0,m}^2$	Variance of level 3 intercepts for outcome m
$\eta_{1,m}^2$	Variance of level 3 impacts for outcome m (cross-district treatment heterogeneity)
$\boldsymbol{\xi}_m$	Coefficient vector of level 3 covariates $\mathbf{V}_{km}$
$u_{0,jkm}$	Level 2/School residual intercepts
$u_{1,jkm}$	Level 2/School residual impacts
$\tau_{0,m}^2$	Variance of level 2 intercepts for outcome m
$\tau_{1,m}^2$	Variance of level 2 impacts for outcome m (cross-school treatment heterogeneity)
$\boldsymbol{\delta}_m$	Coefficient vector of level 2 covariates $\mathbf{X}_{jkm}$
r_{ijkm}	Level 1/Individual intercepts
σ_m^2	Variance of individual/level 1 residuals
$\boldsymbol{\gamma}_m$	Coefficient vector of individual/level 1 covariates $\mathbf{C}_{ijkm}$

Table 3: Latent parameters capturing variation. "Residual" indicates deviations from an overall group mean or expectation.

1.4 Scenario naming convention

We denote the research design by d, followed by the number of levels and randomization level, so "d3.1" is a 3-level design with randomization at level 1. The model is denoted by m, followed by the level and the assumption for the intercepts, either f or r and then the assumption for the treatment impacts, c, f, or r. For example, m3ff2rc means at level 3, we assume fixed intercepts and fixed treatment impacts, and at level 2 we assume random intercepts and constant treatment impacts. The full design and model are specified by concatenating these together, e.g., d2.1_m3fc.

Examples:

- d2.1_m2rr: 2 level, individual assignment, level 2 random intercept and random treatment effect. Corresponds to PowerUP! blocked_i1_2r.

- d3.2_m3ff2rc: 3 level, level 2 assignment, level 3 fixed intercepts and fixed treatment effects, level 2 random intercepts and constant treatment effects. Corresponds to PowerUP! blocked_c2_3f.

Table 5 shows the list of supported scenarios and their corresponding names in PowerUp!

1.5 Control parameters

To calculate power, a user must choose assumed values for some of the latent parameters. However, for certain parameters, the user may instead have more intuition about likely values of functions of these parameters, rather than the parameters themselves. For example, rather than choosing the values of the coefficients for all the level 3 covariates one has (ξ_m), the user would set $R_{3,m}^2$, the amount of level three variation explained by the covariates. These derived parameters, which are functions of unobserved parameters, are listed in Table 6. We refer to these derived parameters as "control" parameters because they are set, or controlled, by the user.

PowerUp!	PUMP	Description
β_{0j}	$\theta_{0,jkm}$	Mean outcome under no treatment for school j in district k
β_{1j}	$\theta_{1,jkm}$	Mean impact for school j in district k
X_{ij}	$\mathbf{C}_{ijkm}$	Individual covariates
β_{2j}	$\boldsymbol{\gamma}_m$	Coefficient vector for individual covariates C_{ijkm}
γ_{00}	$\psi_{0,km}$	Grand mean outcome under no treatment across schools in district k
γ_{10}	$\psi_{1,km}$	Grand mean impact across schools in district k
W_{jk}	$\mathbf{X}_{jkm}$	School covariates
γ_{01k}	$\boldsymbol{\delta}_m$	Coefficient vector for school covariates $\mathbf{X}_{jkm}$
μ_{0j}	$u_{0,jkm}$	School intercepts
μ_{1j}	$u_{1,jkm}$	School impacts
$\tau^2_{2|W}$	$\tau^2_{0,m}$	Variance of school random effects
τ^2_2	$\tau^2_{0,m} + \delta^2_m$	Overall variance of schools
$\tau^2_{T2|W}$	$\tau^2_{1,m}$	Variance of school impacts
ρ_2	ICC_2	Intraclass correlation (unconditional) for level 2
ω_2	ω_2	Ratio of variation of impacts to residuals for level 2
τ_{2T2}	$\boldsymbol{\kappa}^u$	Correlations between school random effects and impacts
ξ_{000}	$\Xi_{0,m}$	Grand mean outcome under no treatment across districts
ξ_{100}	$\Xi_{1,m}$	Grand mean impact across districts
V_k	$\mathbf{V}_{km}$	District covariates
ξ_{001}	$\boldsymbol{\xi}_m$	Coefficient vector for district covariates $\mathbf{V}_{km}$
ζ_{00}	$w_{0,km}$	District intercepts
ζ_{10}	$w_{1,km}$	District impacts
$\tau^2_{3|V}$	$\eta^2_{0,m}$	Variance of district random effects
τ^2_3	$\eta^2_{0,m} + \xi^2_m$	Overall variance of districts
$\tau^2_{T3|V}$	$\eta^2_{1,m}$	Variance of district impacts
τ_{3T3}	$\boldsymbol{\kappa}^w$	Correlations between district random effects and impacts

Table 4: Correspondence of parameters with PowerUp!

We next provide further clarification on these control parameters. To keep the formulae for intraclass correlation coefficients (ICCs) and R^2 terms simple and clear, we assume we only have a single covariate at each level for the purpose of these derivations. We further assume that all covariates have unit variance and are group-mean centered.[1] Group-mean centering means covariates can only explain variation at their level; in practice, raw lower-level covariates could explain variation in higher levels, if they systematically differ by group (e.g., individual student tests might capture the "contextual effect" that could be predicted with school-level average test scores). This again allows for clean formula for how covariates can impact the overall standard errors.

---

[1]A group-mean centered covariate is when you subtract the mean covariate value off for each group in turn. For example, to group-mean center student math tests, you would calculate school-average math test for each school, and then subtract that off of the individual student scores. The school-level mean of a group-mean-centered covariate will always be zero.

PUMP Model	Name in PowerUp!	Other Names
d1.1_m1c	n/a	Simple randomization
d2.1_m2fc	bira2_1c	RICC
d2.1_m2ff	bira2_1f	
d2.1_m2fr or d2.1_m2rr	bira2_1r	FIRC and RIRC
d2.2_m2rc	cra2_2r	
d3.1_m3rr2rr	bira3_1r	
d3.2_m3ff2rc	bcra3_2f	
d3.2_m3fc2rc	n/a	
d3.2_m3rr2rc	bcra3_2r	
d3.3_m3rc2rc	cra3_3r	

Table 5: Scenarios (designs and models) supported in PUMP.

Param	Description
ES_m	average treatment impact in effect size units
$\mathrm{ICC}_{3,m}$	level 3 (district) intraclass correlation
$\omega_{3,m}$	ratio of variation of district average impacts to district intercepts
$R^2_{3,m}$	percent of district variation explained by level 3 (district) covariates $\mathbf{V}_{km}$
$\mathrm{ICC}_{2,m}$	level 2 (school) intraclass correlation
$\omega_{2,m}$	ratio of variation of school average impacts to school intercepts
$R^2_{2,m}$	percent of school variation explained by level 2 (school) covariates $\mathbf{X}_{jkm}$
$R^2_{1,m}$	percent of individual variation explained by level 1 (individual) covariates C_{ijkm}

Table 6: Control parameters.

Using the simplifying assumptions of single, centered, covariates are without real loss of generality. Expressing scenarios in terms of the overall parameters such as the R^2 measures allows for representing our models in this massively simplified way. Thus, in the package itself and in our explanation of the Scenarios in Section 2, we explicitly allow for the models to contain multiple and arbitrary covariates. For using the power formula that derive from the simpler models, researchers simply input total R^2 and ICC values that include the full explanatory power of all covariates for a given level. The results will be equivalent.

Effect size. The first control parameter, the effect size, is defined as the average treatment effect divided by the total control variation:

$$ES_m = \frac{\mathrm{ATE}_m}{\sqrt{Var(Y_{ijkm}(0))}} = \frac{\mathrm{ATE}_m}{\sqrt{(\xi_m^2 + \gamma_m^2) + (\delta_m^2 + \eta_{0,m}^2) + (\tau_{0,m}^2 + \sigma_m^2)}}. \tag{7}$$

Remember that the definition of the ATE changes depending on the model: for three-level models it corresponds to the average district treatment effect $\Xi_{1,m}$, for two-level models the average school treatment effect $\psi_{1,m}$, and for one-level models the average student treatment effect $\theta_{1,m}$.

Note. We can also define the effect size in terms of the level one treatment and control means. We can define the following quantities:

$$\mu_{C,m} = \frac{1}{KJ\bar{n}} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{n=1}^{\bar{n}} Y_{ijkm}(0) \tag{8}$$

$$\mu_{T,m} = \frac{1}{KJ\bar{n}} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{n=1}^{\bar{n}} Y_{ijkm}(1). \tag{9}$$

Then, the effect size is

$$ES_m = \frac{\mu_{T,m} - \mu_{C,m}}{\sqrt{Var(Y_{ijkm}(0))}}. \tag{10}$$

The equivalence between $\mu_{T,m} - \mu_{C,m}$ and ATE_m holds because we assume independence between sample size and any treatment impact heterogeneity (e.g., we don't have a setting where larger schools have a systematically different treatment impact size from smaller schools).

ICC. The ICCs are unconditional Intraclass Correlations, meaning they include the variation explained by covariates. Because we assume all covariates are group-mean centered and have unit variance, we get the pairing structure of terms in the equations below.

$$\text{ICC}_{3,m} = \frac{Var(\psi_{0,km})}{Var(Y_{ijkm}(0))} \qquad = \frac{\xi_m^2 + \eta_{0,m}^2}{(\xi_m^2 + \eta_{0,m}^2) + (\delta_m^2 + \tau_{0,m}^2) + (\gamma_m^2 + \sigma_m^2)} \tag{11}$$

$$\text{ICC}_{2,m} = \frac{Var(\theta_{0,jkm} \mid \psi_{0,km})}{Var(Y_{ijkm}(0))} \qquad = \frac{\delta_m^2 + \tau_{0,m}^2}{(\xi_m^2 + \eta_{0,m}^2) + (\delta_m^2 + \tau_{0,m}^2) + (\gamma_m^2 + \sigma_m^2)}. \tag{12}$$

The squared coefficients γ_m^2, δ_m^2, and ξ_m^2 represent the total predictive power of the covariates at each level. The tighter the coupling of a covariate and outcome, the more the overall outcome variation will be due to that covariate. For the multiple-covariate version, replace those terms with, e.g., $Var(\boldsymbol{\gamma}_m' \mathbf{C}_{ijkm})$. The second term of each pair is due the variation in the intercepts themselves: this is variation not explainable by the covariates in the data.

Impact variation. The quantity ω is the ratio between the variation in average impact of a unit and the variation in the control-side mean of a unit.

$$\omega_{3,m} = \frac{Var(\psi_{1,km})}{Var(\psi_{0,km})} \qquad = \frac{\eta_{1,m}^2}{\xi_m^2 + \eta_{0,m}^2} \tag{13}$$

$$\omega_{2,m} = \frac{Var(\theta_{1,jkm} \mid \psi_{1,km})}{Var(\theta_{0,jkm} \mid \psi_{0,km})} \qquad = \frac{\tau_{1,m}^2}{\delta_m^2 + \tau_{0,m}^2}. \tag{14}$$

Covariate-outcome relationships. The R^2 expressions are the percent of variation at a particular level predicted by covariates. The group-mean centering makes these formulae only involve covariates at the same level as the R^2; this again is a simplification of convenience. The standard error formula for the models in the remainder of the document are general, however, and a user would not need to group-mean center or rescale any covariate in practice.

$$R^2_{3,m} = 1 - \frac{Var(w_{0,km})}{Var(\psi_{0,km})} \qquad\qquad = 1 - \frac{\eta^2_{0,m}}{\xi^2_m + \eta^2_{0,m}} \tag{15}$$

$$R^2_{2,m} = 1 - \frac{Var(u_{0,jkm})}{Var(\theta_{0,jkm} \mid D_{id})} \qquad\qquad = 1 - \frac{\tau^2_{0,m}}{\delta^2_m + \tau^2_{0,m}} \tag{16}$$

$$R^2_{1,m} = 1 - \frac{Var(r_{ijkm})}{Var(Y_{ijkm}(0) \mid S_{id}, D_{id})} \qquad\qquad = 1 - \frac{\sigma^2_m}{\gamma^2_m + \sigma^2_m}. \tag{17}$$

1.6   Power estimation strategy

The same strategy is followed for all designs. First, we lay out a model for our outcomes, Y_{ijkm}. Next, we calculate the standard error of the average treatment effect estimate. When expressing the estimated treatment effect as an effect size, the standard error is given by:

$$Q_m \equiv \text{SE}\left(\widehat{\text{ES}}_m\right) = \text{SE}\left(\frac{\widehat{\text{ATE}}_m}{\sqrt{\text{VAR}_m}}\right) = \frac{1}{\sqrt{\text{VAR}_m}}\text{SE}\left(\widehat{\text{ATE}}_m\right). \tag{18}$$

The standard error formulae are derived assuming there is no treatment variation other than that explicitly captured by the model; in general, if there is additional variation then this could induce heteroskedasticity which could impact power and, in principle, validity. This could come up if different models for the same design are being compared, e.g., for d2.1 designs, if the m2fc model is being compared to the m2fr model, with the first assuming the same impact across sites and the second heterogeneous impacts across sites. Unless impact heterogeneity is substantial, however, the standard error formula will be essentially correct; we recommend ignoring this potential source of approximation error in the course of conducting power analyses.

When analyzing actual data, we would, to estimate Q_m, plug in known values for $\bar{T}$, J, and $\bar{n}$. Any other parameters would be replaced by sample estimates. Then, when testing the mth null hypothesis, $\text{ES}_m = 0$, the test statistic for a t-test is given by

$$t_m \equiv \frac{\widehat{\text{ES}}_m}{\widehat{Q}_m}. \tag{19}$$

When the null is true, t_m follows a t distribution with mean 0 and degrees of freedom df_m, which depends on the design and model.

For power calculations we calculate, given our assumptions on the design and selected model, a reasonable value for Q_m. We can then calculate the power to detect an impact expressed in effect size units.

From the power formulae, we can also calculate MDES and sample size requirements. From Dong and Maynard (2013), in general the MDES can be estimated as

$$\text{MDES} = \text{MT}_{df} \times \frac{\text{SE}}{\sqrt{\text{VAR}}}, \tag{20}$$

where MT_{df} is known as the multiplier and is the sum of two t statistics based on degrees of freedom df. For one-tailed tests, $\text{MT}_{df} = t_\alpha + t_{1-\beta}$ where α is the type I error rate and β is the desired power. For two-tailed tests, $\text{MT}_{df} = t_{\alpha/2} + t_{1-\beta}$. For more details, see Dong and Maynard (2013, page 31) or Bloom (2006, page 22). Manipulating this expression then results in sample size formulae.

A note on effect sizes. In describing the standard error of our estimators in terms of effect size, we need to carefully identify what we mean by an "effect size." We commonly think of an effect size as the size of an impact relative to some reference amount of variation. If the reference amount of variation is different, then the effect size, for the same absolute effect, will also be different. This concern of what the denominator should be can create some tension regarding some of the power formula.

In a 2 level design, "one might define the effect size in terms of a standard deviation based on the variance between level 1 units, σ_1^2, the variance between level 2 units, σ_2^2, or the total variance $\sigma_1^2 + \sigma_2^2$" (Spybrook et al., 2014a). We follow the definition of using the total variance over all levels in the denominator, which is made explicit in Equation 7.

This is most obviously a concern with fixed effect regression. In particular, relative to overall variation, if we increase the ICC at level 2 or level 3, then there is less variation (relative to the reference variation) in level 1; thus an increased ICC will increase power for fixed effect regression. This is simply the realized gains of a blocked experiment. If the effect size is calculated relative to within-group variation, however, this gain would not be seen.

Defining effect size, i.e., the reference variation, is not consistent across all power calculators, as is discussed in, for example, Spybrook et al. (2014b). It can even differ within the same calculator; the PowerUp! framework, for example, includes between cluster variation for random effects models and does not for fixed effect equivalents. Similarly, OD Plus defines effect size differently for cluster randomized designs (where it includes between-cluster variation) and randomized block designs (where it uses only within-block residual variation); see Tables 2 and 3 in Spybrook et al. (2014b). We try to keep the models aligned by always using the same reference variation of overall or total variation.

An alternate approach for aligning fixed vs. random effect models is to include variance explained by the fixed effects in the R^2 measure; for example, if there is substantial district variation, this can be captured with a higher level two R^2, if we consider the district ids of schools as level two covariates. To make our formula more directly comparable, we do not take this route, but one can use our package and obtain the same results by selecting an appropriate R^2 and then setting the next higher level ICC to 0.

Required user-provided values. In the above sections, we have defined a large number of quantities, including latent, observed, and control quantities. We summarize all of the quantities that the user needs to specify in order to estimate power:

- Minimum detectable effect size (MDES) for each outcome.

- Sample sizes at each level $(\bar{n}, J, K)$.

- ICC values at each level for each outcome.

- R^2 values at each level for each outcome.

- ω values at each level for each outcome.

- Number of covariates at each level.

These parameters are also summarized in Table 3 in the main text.

2 Scenarios

2.1 d1.1 designs: 1 level, randomization at level 1

This is the classic individually randomized experiment where we allocate some fraction of a single set of units to treatment.

The randomization scheme is simple random sampling:

```
T.x <- randomizr::simple_ra(N = nbar, prob = Tbar)
```

2.1.1 Constant effects (d1.1_m1c)

PowerUp! name. Not applicable.

Design. 1-level design, randomization at level 1

Model. constant intercepts, constant treatment effects, no school or district covariates.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{1,m}T + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm}. \tag{21}$$

Standard error. The standard error of the treatment effect estimate is:

$$Q_m = \sqrt{\frac{(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})\bar{n}}}. \tag{22}$$

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = \bar{n} - g_{1,m} - 1. \tag{23}$$

Sample size formula. The sample size formulae are:

$$\bar{n} = \left(\frac{\text{MT}_{df}}{\text{MDES}}\right)^2 \left(\frac{1 - R_{1,m}^2}{\bar{T}(1 - \bar{T})}\right). \tag{24}$$

Code syntax. The R model is:

```
Yobs ~ 1 + T.x + C.ijk
```

2.2   d2.1 designs: 2 levels, randomization at level 1

This section of designs comprise what are usually referred to as *multisite experiments*. In a multisite experiment, we have a collection of sites (here, schools) and are able to randomize the individuals within each site into treatment and control. This allows for estimating an average impact for each site, in principle. That being said, we are usually interested in estimating some overall summary of impacts across all our sites. These are also called blocked experiments, especially if the sites are viewed as fixed.

Critically, there are four different estimands we might consider: the average impact for persons vs. impact for sites, and the average impact of the sample we have (often referred to as finite-population) vs. the average impact of the population where the sample came from (often referred to as the super-population). When sites are equal sized, a common assumption for power calculations, the site and person average will be the same. We therefore ignore it here. For finite- vs. super-population, we have to be more careful. Some estimation strategies target a finite-population estimand. In this document, the ones that do are d2.1_m2fc and d2.1_m2ff. The d2.1_m2fc estimation strategy does because it assumes a constant treatment impact; given this assumption, there is no uncertainty due to the sample itself as all samples have the same average impact by assumption. The d2.1_m2ff estimation strategy allows each school to have an individually estimated impact, but due to using fixed effects rather than random, it is evaluating the sample at hand. See Miratrix et al. (2021) for further, in-depth, discussion. Estimators that target the super-population need to take any uncertainty of the sample being representative of the super-population into account. Here, the one that does this is d2.1_m2fr, with a model of each school having an average impact drawn from some random distribution.

Regardless of the model used to analyze these data, the randomization scheme is the same. It is simple random sampling within each school, with proportion $\bar{T}$ units assigned to treatment in each school. In R, we could randomize this way as so:

```
T.x <- randomizr::block_ra( S.id, prob = Tbar )
```

2.2.1   Constant effects (d2.1_m2fc, or RICC)

PowerUp! name. bira2_1c

Design. 2-level design, randomization at level 1 (blocked).

Model. fixed intercepts, constant treatment effect, no school covariates.

When we assume constant effects, each school has its own fixed intercept for the control outcome, and the treatment effect is modeled as constant across schools. We can also call this a fixed effects, constant treatment coefficient model, or RICC (Miratrix et al., 2021). This model allows some schools to have higher average outcomes than others (allowed for with the fixed effects), but assumes the treatment impact is the same.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \psi_{1,m}T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm} \tag{25}$$

$$\theta_{0,jkm} = \psi_{0,km} + u_{0,jkm}.$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \psi_{1,m}T_{ijk} + \psi_{0,km} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + u_{0,jkm} + r_{ijkm}, \tag{26}$$

and distributions:

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right). \tag{27}$$

Standard error. The standard error is given by:

$$Q_m = \sqrt{\frac{(1 - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\bar{n}}}. \tag{28}$$

See below for important details on this specific formula, and how it differs from PowerUp!

Degrees of freedom. The degrees of freedom are given by:

$$\mathrm{df}_m = J\bar{n} - g_{1,m} - J - 1. \tag{29}$$

Parameter assumptions. The constant effects model means that we assume no treatment variation across our sites, i.e.,:

- $\omega_{2,m} = 0$.

PowerUp! Differences. PowerUp! assumes there is no $\mathrm{ICC}_{2,m}$ term (i.e., it assumes $\mathrm{ICC}_{2,m} = 0$) while we allow for it. This can be viewed as within (PowerUp!) vs. overall (this work) effect size metrics. We discuss this more next.

Sample size formulae. The sample size formulae are:

$$J = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{n}\bar{T}(1 - \bar{T})}\right) \tag{30}$$

$$\bar{n} = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}{J\bar{T}(1 - \bar{T})}\right). \tag{31}$$

Code syntax. The R model is:

```
Yobs ~ 0 + T.x + C.ijk + S.id
```

The 0 coupled with the `S.id` gives each school an independently estimated intercept rather than selecting one school as a reference category; the overall mean parameter $\psi_{0,km}$ is thus left implicit in the R call. One could estimate it by averaging the individual fixed effect coefficients for the dummy variables generated by `S.id`.

15

Remark on effect sizes. The standard error of the treatment effect estimate *not* in effect size units is (this taken from the PowerUp! documentation):

$$SE(\widehat{ATE}_m) = \sqrt{\frac{1}{\bar{T}(1-\bar{T})J\bar{n}}} \cdot \sigma_m. \tag{32}$$

To convert this to an effect size, we need to scale by overall variation. Unfortunately, under a fixed effect model, there is no natural way to express this as we have not parameterized how the individual site intercepts, the $\delta_{0,jkm}$, vary. PowerUp! therefore indexes by within group variation, which is

$$Var(Y_{ijkm}(0)|S_{id}) = \frac{\sigma_m^2}{1-R_{1,m}^2},$$

using the formula for $R_{1,m}^2$ of

$$R_{1,m}^2 = 1 - \frac{\sigma_m^2}{Var(Y_{ijkm}(0)|S_{id})}.$$

The above expression captures the predictive power of our individual-level covariates on the outcomes within a given school. If we divide the above $SE(\widehat{ATE}_m)$ formula by $\sqrt{\sigma_m^2/(1-R_{1,m}^2)}$ we get the reported standard error formula for Q_m of

$$\tilde{Q}_m = \sqrt{\frac{1-R_{1,m}^2}{\bar{T}(1-\bar{T})J\bar{n}}},$$

with the tilde denoting that these effect size units are in terms of within-school variation. Equivalently, this formula assumes the blocks (schools) are all homogeneous, which both goes counter to the design principles of blocking and also is known to generally not hold when evaluating schools. If we want the more classic effect size indexed by overall variation, we need to go further.

Assume we have an $ICC_{2,m}$, an assumed measure of how much overall (control-side) variation is at the school level:

$$ICC_{2,m} = 1 - \frac{Var(Y_{ijkm}(0)|S_{id}))}{Var(Y_{ijkm}(0))}.$$

This ICC is even defined for a finite sample, if we view the above as comparing the empirical (pooled) within-group variation to full variation. Rearranging this gives $Var(Y_{ijkm}(0)) = Var(Y_{ijkm}(0)|S_{id})/(1-ICC_{2,m})$.

We can then plug this and the $R_{1,m}^2$ formula together to get

$$Var(Y_{ijkm}(0)) = \frac{\sigma_m^2}{1-R_{1,m}^2} \cdot \frac{1}{1-ICC_{2,m}}.$$

If we use this expression to scale our SE formula, we obtain our formula listed above.

2.2.2 Fixed effects (d2.1_m2ff)

PowerUp! name. bira2_1f

Design. 2-level design, randomization at level 1 (blocked).

Model. fixed intercepts, fixed treatment effects, no school covariates.

The constant effects model assumes treatment is the same for each block. If it is not, and the blocks are different sizes or have different proportions of units treated, the constant effects estimator is precision-weighted and can thus be biased. Some may instead choose to allow each school to have its own estimated impact, with a second averaging step where we calculate an overall site-average of the site specific impact estimates.

We do this by interacting our site fixed effects with treatment. Now each school has its own fixed intercept for the control outcome, and each school also has its own fixed coefficient for the treatment effect. We can also call this a fixed effects with interactions model (Miratrix et al., 2021).

In practice, the power calculations for this model will be the same as for constant effects, unless we allow for block size variation or variable proportion treated.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{1,jkm}T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm} \tag{33}$$

$$\theta_{0,jkm} = \psi_{0,km} + u_{0,jkm}$$
$$\theta_{1,jkm} = \psi_{1,km} + u_{1,jkm},$$

and distributions:

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right). \tag{34}$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \left(\psi_{1,km} + u_{1,jkm}\right)T_{ijk} + \psi_{0,km} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + u_{0,jkm} + r_{ijkm}. \tag{35}$$

Standard error. The standard error of the treatment effect estimate (and therefore the sample size formula) are all the same as in the constant effects model, i.e. for SE we have:

$$Q_m = \sqrt{\frac{(1 - \text{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\bar{n}}}. \tag{36}$$

Degrees of freedom. However, the degrees of freedom are different due to the additional interaction terms we need to estimate:

$$\text{df}_m = J\bar{n} - g_{1,m} - 2J. \tag{37}$$

PowerUp! Differences. Just as with the constant model, PowerUp! assumes there is no $\text{ICC}_{2,m}$ term while we allow for it. This can be viewed as within (PowerUp!) vs. overall (this work) effect size metrics; see the discussion at the end of the constant effects model, above, for more detail.

Sample size formula. The sample size formulae are:

$$J = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{n}\bar{T}(1 - \bar{T})}\right) \tag{38}$$

$$\bar{n} = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}{J\bar{T}(1 - \bar{T})}\right). \tag{39}$$

Code syntax. The R model is:

```
Yobs ~ 0 + T.x:S.id - T.x + C.ijk
```

The '`- T.x`' term forces a separate impact estimate for each school (which come from the interaction term of `T.x:S.id`). Neither $\psi_{0,km}$ or $\psi_{1,km}$ are explicitly estimated; you would obtain estimates by averaging the corresponding individual estimates. If schools are different sizes, different weighting choices are possible here (see, again, Miratrix et al. (2021)). Standard error estimates are also obtained by aggregating the squared standard error estimates for the individual schools.

2.2.3   Random effects (d2.1_m2fr and d2.1_m2rr, or FIRC and RIRC)

PowerUp! name. bira2_1r

Design. 2-level design, randomization at level 1 (blocked).

Model. random intercepts, random treatment effect, school covariates for intercept. PowerUp! also includes interaction terms of treatment and school covariates to allow for modeling treatment effect heterogeneity; we do not include this.

If we are interested in generalizing from our sample to a super-population, we may wish to view the sample of schools themselves as representative of something larger. Then, if some schools have different average impacts than other schools, we have to account for the possibility that our sample of schools has an overall average impact different from the target population. We can account for this additional uncertainty with a random effects model that has a random effect for the school-level average impacts.

The classic random effects model gives each school both a random intercept for the control average outcome (the intercept), and a random coefficient for the treatment effect. This is also known as the RIRC model: random intercept, random coefficient. Recently, researchers also use a variant of this model, the Fixed Intercept, Random Coefficient (FIRC) model to account for concerns such as varying proportions of units treated in different schools.[2] For power calculations, FIRC and RIRC have the same performance (they are also similar in practice; see Miratrix et al. (2021)).

For RIRC, the model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{1,jkm}T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm} \tag{40}$$

$$\theta_{0,jkm} = \psi_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + u_{0,jkm}$$

$$\theta_{1,jkm} = \psi_{1,km} + u_{1,jkm},$$

and random effect and residual distributions of:

$$\begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{0,m}^2 & \kappa_{mm}^u \tau_{0,m}\tau_{1,m} \\ \kappa_{mm}^u \tau_{1,m}\tau_{0,m} & \tau_{1,m}^2 \end{pmatrix} \right) \tag{41}$$

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

For FIRC, we only have the random effects model on the $u_{1,jkm}$, and have fixed effects for the $u_{0,jkm}$. For RIRC, we assume bivariate Normal effects with variances $\tau_{0,m}^2$ and $\tau_{1,m}^2$ and correlation κ_{mm}^u. The correlation structure κ_{mm}^u does not heavily impact the distribution of the final test statistic.

Reduced form. The reduced form is:

$$Y_{ijkm} = (\psi_{1,km} + u_{1,jkm})T_{ijk} + \psi_{0,km} \tag{42}$$

$$+ \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + u_{0,jkm} + r_{ijkm}.$$

---

[2]For fitting this model in R, see https://cares-blog.gse.harvard.edu/post/fitting-firc/.

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})J\bar{n}}}. \tag{43}$$

Note that this formula is simply the formula for d2.1_m2fc with an additional term of $\text{ICC}_{2,m}\omega_{2,m}/J$. The new term captures the additional uncertainty from extrapolating from our sample to the super-population. Q_m with this model, therefore, will be larger than the prior models to the extent that the schools differ in terms of their impact variation (the $\text{ICC}_{2,m}\omega_{2,m}$ term is simply the variation in the random impact terms scaled by our overall variation).

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = J - g_{1,m} - 1. \tag{44}$$

PowerUp! Differences. PowerUp! assumes that school and district covariates also influence the treatment impact:

$$\theta_{1,jkm} = \psi_{1,km} + \sum_{r=1}^{g_{2,m}} \phi_{mr}X_{jkmr} + u_{1,jkm}.$$

We do not make this assumption. The result of this is that we assume, in their notation, that $R^2_{2T} = 0$, where R^2_{2T} is the percent of treatment variation explained by level 2 covariates. This assumption affects the first term in the standard error formula.

Sample size formula. The sample size formulae are:

$$J = \left(\frac{\text{MT}_{df}}{\text{MDES}}\right)^2 \left(\text{ICC}_{2,m}\omega_{2,m} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})\bar{n}}\right) \tag{45}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})\left(J\left(\frac{\text{MDES}}{\text{MT}_{df}}\right)^2 - \text{ICC}_{2,m}\omega_{2,m}\right)}. \tag{46}$$

Code syntax. The R model is, for RIRC:

```
Yobs ~ 1 + T.x + X.jk + C.ijk + (1 + T.x | S.id)
```

For FIRC it is:

```
Yobs ~ 0 + T.x + X.jk + C.ijk + S.id + (0 + T.x | S.id)
```

2.3 d2.2 designs: 2 levels, randomization at level 2

These are commonly called cluster randomized experiments, with the schools being the clusters.

The randomization scheme is a simple random sample of $J\bar{T}$ schools assigned to treatment:

```
T.x <- randomizr::cluster_ra( S.id, prob = Tbar )
```

2.3.1 Random effects (d2.2_m2rc)

PowerUp! name. cra2_2r

Design. 2-level design, randomization at level 2 (clusters).

Model. random intercepts, constant treatment effect for all schools, school covariates for intercept. The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm} \tag{47}$$

$$\theta_{0,jkm} = \psi_{0,km} + \theta_{1,m}T_{jk} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + u_{0,jkm},$$

and distributions:

$$u_{0,jkm} \sim N\left(0, \tau_{0,m}^2\right) \tag{48}$$

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \theta_{1,m}T_{jk} + \psi_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + u_{0,jkm} + r_{ijkm}. \tag{49}$$

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1-\bar{T})J} + \frac{(1 - \text{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})J\bar{n}}}. \tag{50}$$

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = J - g_{1,m} - 2. \tag{51}$$

Parameter assumptions. The constant effects model means that we assume no treatment variation across our sites, i.e.:

- $\omega_{2,m} = 0$.

Sample size formula. The sample size formulae are:

$$J = \left(\frac{\text{MT}_{df}}{\text{MDES}}\right)^2 \left(\frac{\bar{n}\text{ICC}_{2,m}(1 - R_{2,m}^2) + (1 - \text{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})\bar{n}}\right) \tag{52}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\left(\frac{\text{MDES}}{\text{MT}_{df}}\right)^2 - \text{ICC}_{2,m}(1 - R_{2,m}^2).} \tag{53}$$

Code syntax. The R model is:

```
Yobs ~ 1 + T.x + X.jk + C.ijk + (1 | S.id)
```

2.4   d3.1 designs: 3 levels, randomization at level 1

In these designs we have schools nested in districts, and students nested in schools. The only difference here, as compared to blocked individual randomization with two levels, is the third level of district. Since we are randomizing at the student level, this will only impact how we think about where variation is in terms of our effect size units.

In this context, if we are interested in the finite-sample impacts, other than for calculating our reference variation for effect sizes, the districts do not matter. We can simply use the prior two level fixed effect designs if we lump district variation into the $\text{ICC}_{2,m}$ terms. In particular, one could use d2.1_m2ff or d2.1_m2fc for the three level case by just entering $\text{ICC}_{2,m} + \text{ICC}_{3,m}$ in for $\text{ICC}_{2,m}$. In fact, we cannot have district random or fixed effects given school-level fixed effects due to collinearity.

For superpopulation impacts, we can either view the districts and schools as random draws, or the districts as fixed, with schools randomly drawn within them. This gives the two models discussed below.

Regardless, the randomization scheme is: simple random sampling occurs within each school, with proportion $\bar{T}$ units assigned to treatment in each school.

```
T.x <- randomizr::block_ra( S.id, prob = Tbar )
```

2.4.1   Random effects (d3.1_m3rr2rr)

PowerUp! name. bira3_1r

Design. 3-level design, randomization at level 1 (blocked).

Model. random intercepts for district, random treatment effects for district, random intercepts for school, random effects for schools, school and district covariates for intercepts. PowerUp! also allows for school and district covariates for cross-site impact heterogeneity.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{54}$$

$$\theta_{0,jkm} = \psi_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm}$$

$$\theta_{1,jkm} = \psi_{1,km} + u_{1,jkm}$$

$$\psi_{0,km} = \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km}$$

$$\psi_{1,km} = \Xi_{1,m} + w_{1,km},$$

and distributions:

$$
\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta_{0,m}^2 & \kappa_{mm}^w \eta_{0,m}\eta_{1,m} \\ \kappa_{mm}^w \eta_{1,m}\eta_{0,m} & \eta_{1,m}^2 \end{pmatrix} \right)
$$

$$
\begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{0,m}^2 & \kappa_{mm}^u \tau_{0,m}\tau_{1,m} \\ \kappa_{mm}^u \tau_{1,m}\tau_{0,m} & \tau_{1,m}^2 \end{pmatrix} \right) \tag{55}
$$

$$
r_{ijkm} \sim N\left(0, \sigma_m^2\right).
$$

Reduced form. The reduced form is:

$$
Y_{ijkm} = \left(\Xi_{1,jkm} + w_{1,km} + u_{1,jkm}\right)T_{ijk} + \Xi_{0,km} \tag{56}
$$
$$
+ \sum_{s=1}^{g_{3,m}} \xi_{ms}V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp}
$$
$$
+ w_{0,km} + u_{0,jkm} + r_{ijkm}.
$$

Standard error. The standard error of the treatment effect estimate is given by:

$$
Q_m = \sqrt{\frac{\mathrm{ICC}_{3,m}\omega_{3,m}}{K} + \frac{\mathrm{ICC}_{2,m}\omega_{2,m}}{JK} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{57}
$$

Degrees of freedom. The degrees of freedom are given by:

$$
\mathrm{df}_m = K - 1. \tag{58}
$$

This is a very conservative degrees of freedom.

PowerUp! Differences. Similar to the two-level blocked model, in PowerUp! they further assume that school and district covariates also influence the treatment impact:

$$
\theta_{1,jkm} = \psi_{1,km} + \sum_{r=1}^{g_{2,m}} \phi_{mr}X_{jkmr} + u_{1,jkm}
$$
$$
\psi_{1,km} = \Xi_{1,m} + \sum_{s=1}^{g_{3,m}} \zeta_{mr}V_{kms} + w_{1,km},
$$

but we do not make this assumption. In PowerUp! terms, we assume $R_{3T}^2 = 0$ and $R_{2T}^2 = 0$. The interaction terms change the degrees of freedom from $\mathrm{df}_m = K - g_{3,m} - 1$ to $\mathrm{df}_m = K - 1$.

Sample size formula. The sample size formulae are:

$$
K = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\mathrm{ICC}_{3,m}\omega_{3,m} + \frac{\mathrm{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{59}
$$

$$
J = \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2) + \bar{T}(1-\bar{T})\bar{n}\mathrm{ICC}_{2,m}\omega_{2,m}}{\bar{T}(1-\bar{T})\bar{n}\left(K\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - \mathrm{ICC}_{3,m}\omega_{3,m}\right)} \tag{60}
$$

$$
\bar{n} = \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})\left(JK\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - J\mathrm{ICC}_{3,m}\omega_{3,m} - \mathrm{ICC}_{2,m}\omega_{2,m}\right)}. \tag{61}
$$

24

Code syntax. The R model is:

```
Yobs ~ 1 + T.x + V.k + X.jk + C.ijk + (1 + T.x | S.id) + (1 + T.x | D.id)
```

2.4.2 Fixed district effects (d3.1_m3ff2rr)

PowerUp! name. None.

Design. 3-level design, randomization at level 1 (blocked).

Model. fixed intercepts and treatment effects for district, random intercepts and treatment effects for schools, school covariates for intercepts, individual covariates.

The model for estimating impacts on outcome m is a mild tweak of the prior model–we just replace the random terms at level 3 with fixed effects:

$$Y_{ijkm} = \theta_{1,jkm}T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm} \tag{62}$$

$$\theta_{0,jkm} = \psi_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + u_{0,jkm}$$

$$\theta_{1,jkm} = \psi_{1,km} + u_{1,jkm}$$

$$\psi_{0,km} = \Xi_{0,m} + w_{0,km}$$

$$\psi_{1,km} = \Xi_{1,m} + w_{1,km},$$

and distributions:

$$\begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{0,m}^2 & \kappa_{mm}^u \tau_{0,m}\tau_{1,m} \\ \kappa_{mm}^u \tau_{1,m}\tau_{0,m} & \tau_{1,m}^2 \end{pmatrix} \right) \tag{63}$$

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \left(\Xi_{1,jkm} + w_{1,km} + u_{1,jkm}\right)T_{ijk} + \Xi_{0,km} \tag{64}$$

$$+ \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp}$$

$$+ w_{0,km} + u_{0,jkm} + r_{ijkm}.$$

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}\omega_{2,m}}{JK} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{65}$$

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = JK - K - 1. \tag{66}$$

Sample size formula. The sample size formulae are:

$$K = \left(\frac{\text{MT}_{df}}{\text{MDES}}\right)^2 \left(\frac{\text{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\bar{n}}\right) \tag{67}$$

$$J = \left(\frac{\text{MT}_{df}}{\text{MDES}}\right)^2 \left(\frac{\text{ICC}_{2,m}\omega_{2,m}}{K} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})K\bar{n}}\right) \tag{68}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})\left(JK\left(\frac{\text{MDES}}{\text{MT}_{df}}\right)^2 - \text{ICC}_{2,m}\omega_{2,m}\right)}. \tag{69}$$

Code syntax. The R model is:

```
Yobs ~ 0 + T.x:D.id - T.x + X.jk + C.ijk + D.id + (1 + T.x | S.id)
```

As with the other models with fixed effects interacted with treatment, the individual district-level impact estimates would then be averaged to get the overall impact estimate.

2.5 d3.2 designs: 3 levels, randomization at level 2

These are commonly called blocked, cluster-randomized experiments. You find these if, for example, schools are randomized within a set of districts, or teachers are randomized within a set of schools (with students as outcomes in both cases).

The randomization scheme is: simple random sampling occurs within each district, with $J\bar{T}$ schools assigned to treatment in each district. In R we have:

```
T.x <- randomizr::block_and_cluster_ra( blocks = D.id, clusters = S.id, prob = Tbar )
```

2.5.1 Constant effects (d3.2_m3fc2rc)

PowerUp! name. None.

Design. 3-level design, randomization at level 2 (blocked cluster).

Model. fixed intercepts for districts, constant treatment effect for districts, random intercepts for schools, constant effects for schools within a district, school covariates for intercept.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{70}$$

$$\theta_{0,jkm} = \psi_{0,km} + \Xi_{1,m} T_{jk} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm}$$

$$\psi_{0,km} = \Xi_{0,m} + w_{0,km},$$

and distributions:

$$u_{0,jkm} \sim N\left(0, \tau_{0,m}^2\right) \tag{71}$$
$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \Xi_{1,m} T_{jk} + \Xi_{0,m} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \tag{72}$$

$$+ w_{0,km} + u_{0,jkm} + r_{ijkm}.$$

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1 - \bar{T})JK} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})JK\bar{n}}}. \tag{73}$$

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = K(J - 1) - g_{2,m}. \tag{74}$$

Parameter assumptions. This model assumes no variation of impacts between schools, and no variation at the district level:

- $\omega_{2,m} = 0$.

- $\omega_{3,m} = 0$.

Sample size formula. The sample size formulae are:

$$K = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{\mathrm{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1 - \bar{T})J} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})J\bar{n}}\right) \tag{75}$$

$$J = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{\mathrm{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1 - \bar{T})K} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})K\bar{n}}\right) \tag{76}$$

$$\bar{n} = \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})JK\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - \mathrm{ICC}_{2,m}(1 - R^2_{2,m})}. \tag{77}$$

Code syntax. The R model is:

```
Yobs ~ 0 + T.x + X.jk + C.ijk + D.id + (1 | S.id)
```

28

2.5.2 District fixed effects interacted with treatment (d3.2_m3ff2rc)

PowerUp! name. bcra3_2f

Design. 3-level design, randomization at level 2 (blocked cluster).

Model. fixed intercepts for districts, fixed treatment effects for districts, random intercepts for schools, constant effects for schools within a district, school covariates for intercept.

This model allows for treatment variation at the district level, resulting in K district-specific impact estimates. The model is given by:

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{78}$$

$$\theta_{0,jkm} = \psi_{0,km} + \psi_{1,km} T_{jk} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm}$$

$$\psi_{0,km} = \Xi_{0,m} + w_{0,km}$$

$$\psi_{1,km} = \Xi_{1,m} + w_{1,km},$$

and distributions:

$$u_{0,jkm} \sim N\left(0, \tau_{0,m}^2\right) \tag{79}$$

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

Note that both $w_{0,km}$ and $w_{1,km}$ are fixed effects with no distribution placed on them. The random effect version of this model comes next.

Reduced form. The reduced form is:

$$Y_{ijkm} = \left(\Xi_{1,m} + w_{1,km}\right) T_{jk} + \Xi_{0,m} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \tag{80}$$

$$+ w_{0,km} + u_{0,jkm} + r_{ijkm}.$$

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1 - \bar{T})JK} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})JK\bar{n}}}. \tag{81}$$

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = K(J - 2) - g_{2,m}. \tag{82}$$

Parameter assumptions. This model assumes no variation of impacts between schools within a district:

- $\omega_{2,m} = 0.$

PowerUp! Differences. The PowerUp! formula assumes $\text{ICC}_{3,m} = 0$. This assumption is equivalent to measuring effect size relative to within-district variation as compared to total variation.

Sample size formula. The sample size formulae are:

$$K = \left(\frac{\text{MT}_{df}}{\text{MDES}} \right)^2 \left(\frac{\text{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1 - \bar{T})J} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})J\bar{n}} \right) \tag{83}$$

$$J = \frac{\bar{n}\text{ICC}_{2,m}(1 - R^2_{2,m}) + (1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1 - \bar{T})K \left(\frac{\text{MDES}}{\text{MT}_{df}} \right)^2} \tag{84}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})JK \left(\frac{\text{MDES}}{\text{MT}_{df}} \right)^2 - \text{ICC}_{2,m}(1 - R^2_{2,m})}. \tag{85}$$

Code syntax. The R model is:

```
Yobs ~ 0 + T.x * D.id - T.x + X.jk + C.ijk + (1 | S.id)
```

The overall treatment effect is calculated as the average of the T.x interaction terms. The standard error is calculated as the square root of a weighted sum of the district-specific standard errors squared. When districts are unequal in size, one has different choices as to how to weight these averages.

2.5.3 Random effects (d3.2_m3rr2rc)

PowerUp! name. bcra3_2r

Design. 3-level design, randomization at level 2 (blocked cluster).

Model. random intercepts for districts, random treatment effect for districts, random intercepts for schools, constant effects for schools within a district, school and district covariates for intercept. PowerUp! also allows for district covariates for treatment effects.

The model for estimating impacts on outcome m is the same as that for m3ff2rc (Equation 78), except we place a distribution on the $w_{0,km}, w_{1,km}$ pairs and allow for covariates at level three. This gives:

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{86}$$

$$\theta_{0,jkm} = \psi_{0,km} + \psi_{1,km} T_{jk} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm}$$

$$\psi_{0,km} = \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km}$$

$$\psi_{1,km} = \Xi_{1,m} + w_{1,km},$$

and distributions:

$$u_{0,jkm} \sim N\left(0, \tau_{0,m}^2\right) \tag{87}$$

$$\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta_{0,m}^2 & \kappa_{mm}^w \eta_{0,m} \eta_{1,m} \\ \kappa_{mm}^w \eta_{1,m} \eta_{0,m} & \eta_{1,m}^2 \end{pmatrix} \right)$$

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \left(\Xi_{1,m} + w_{1,km}\right) T_{jk} + \Xi_{0,m} \tag{88}$$

$$+ \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp}$$

$$+ w_{0,km} + u_{0,jkm} + r_{ijkm}.$$

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{3,m}\omega_{3,m}}{K} + \frac{\text{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1 - \bar{T})JK} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})JK\bar{n}}}. \tag{89}$$

Degrees of freedom. The degrees of freedom are given by:

$$\text{df}_m = K - 1. \tag{90}$$

Parameter assumptions. This model assumes no variation of impacts between schools within a district:

- $\omega_{2,m} = 0$.

PowerUp! Differences.   Similar to other blocked models model, PowerUp! further assumes that district covariates also influence the treatment impact:

$$\psi_{1,km} = \Xi_{1,m} + \sum_{s=1}^{g_{3,m}} \zeta_{mr} V_{kms} + w_{1,km}.$$

We do not make this assumption. In PowerUp! terms, we assume $R_{3T}^2 = 0$. This also impacts our degrees of freedom formula, which is $\mathrm{df}_m = K - 1$ instead of $\mathrm{df}_m = K - g_{3,m} - 1$.

Sample size formula.   The sample size formulae are:

$$K = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\mathrm{ICC}_{3,m}\omega_3 + \frac{\mathrm{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1 - \bar{T})J} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\bar{n}}\right) \tag{91}$$

$$J = \frac{\bar{n}\mathrm{ICC}_{2,m}(1 - R_{2,m}^2) + (1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{n}\bar{T}(1 - \bar{T})\left(K\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - \mathrm{ICC}_{3,m}\omega_3\right)} \tag{92}$$

$$\bar{n} = \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\left(K\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - \mathrm{ICC}_{3,m}\omega_{3,m}\right) - \mathrm{ICC}_{2,m}(1 - R_{2,m}^2)}. \tag{93}$$

Code syntax.   The R model is:

```
Yobs ~ 1 + T.x + V.k + X.jk + C.ijk + (1 | S.id) + (1 + T.x | D.id)
```

2.6 d3.3 designs: 3 levels, randomization at level 3

These designs have randomization at the top level. They are cluster randomized, but we can model the nesting structure within cluster. Because randomization is at the top level, we cannot have random or fixed effects for district impacts or school impacts, restricting the models possible for this design.

The randomization scheme is: simple random sampling occurs across districts, with $K\bar{T}$ districts assigned to treatment.

```
T.x <- randomizr::cluster_ra( D.id, prob = Tbar )
```

2.6.1 Random effects (d3.3_m3rc2rc)

PowerUp! name. cra3_3r

Design. 3-level design, randomization at level 3 (cluster).

Model. random intercepts for districts, constant treatment effects for districts, random intercepts for schools, constant treatment effects for schools, school and district covariates for intercept.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp} + r_{ijkm} \tag{94}$$

$$\theta_{0,jkm} = \psi_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + u_{0,jkm}$$

$$\psi_{0,km} = \Xi_{0,m} + \Xi_{1,m}T_k + \sum_{s=1}^{g_{3,m}} \xi_{ms}V_{kms} + w_{0,km},$$

and distributions:

$$u_{0,jkm} \sim N\left(0, \tau_{0,m}^2\right) \tag{95}$$
$$w_{0,jkm} \sim N\left(0, \eta_{0,m}^2\right)$$
$$r_{ijkm} \sim N\left(0, \sigma_m^2\right).$$

Reduced form. The reduced form is:

$$Y_{ijkm} = \Xi_{1,m}T_k + \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms}V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr}X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp}C_{ijkmp}. \tag{96}$$

$$+ w_{0,km} + u_{0,jkm} + r_{ijkm}.$$

Standard error. The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{3,m}(1 - R_{3,m}^2)}{\bar{T}(1-\bar{T})K} + \frac{\text{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1-\bar{T})JK} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{97}$$

Degrees of freedom. The degrees of freedom are given by:

$$\mathrm{df}_m = K - g_{3,m} - 2. \tag{98}$$

Parameter assumptions. The constant effects model means that we assume no treatment variation across our sites, i.e.:

- $\omega_{2,m} = 0$.

- $\omega_{3,m} = 0$.

Sample size formula. The sample size formulae are:

$$K = \left(\frac{\mathrm{MT}_{df}}{\mathrm{MDES}}\right)^2 \left(\frac{\mathrm{ICC}_{3,m}(1 - R_{3,m}^2)}{\bar{T}(1 - \bar{T})} + \frac{\mathrm{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1 - \bar{T})J} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})J\bar{n}}\right)$$

$$\tag{99}$$

$$J = \frac{\bar{n}\mathrm{ICC}_{2,m}(1 - R_{2,m}^2) + (1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{n}\left(\bar{T}(1 - \bar{T})K\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - \mathrm{ICC}_{3,m}(1 - R_{3,m}^2)\right)} \tag{100}$$

$$\bar{n} = \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1 - \bar{T})JK\left(\frac{\mathrm{MDES}}{\mathrm{MT}_{df}}\right)^2 - J\mathrm{ICC}_{3,m}(1 - R_{3,m}^2) - \mathrm{ICC}_{2,m}(1 - R_{2,m}^2)}. \tag{101}$$

Code syntax. The R model is:

```
Yobs ~ 1 + T.x + V.k + X.jk + C.ijk + (1 | S.id) + (1 | D.id)
```

3 The data generating process

We now discuss the assumed data generating process (DGP), indexed by parameters directly tied to the structural equations we use. For simplicity, we outline a DGP that contains at most one covariate per level. In particular, we generate, for each outcome, a single covariate with zero mean and unit variance at each level, and then follow the full random effects model (3rr2rr). The data generating process is done via a series of stages that are detailed below:

1. Generate the K districts by first generating covariates, then random effects.

2. Generate the $K \times J$ schools in a similar manner to districts.

3. Generate the $K \times J \times \bar{n}$ students, including the student-level potential outcomes.

4. Randomize the students according to the specified design, and calculate final observed outcomes.

To use the data generator, we need to have values for all the model parameters. This involves translating the user parameters to these model parameters; we discuss how we do this in Section 4.

For two- or one-level data, we just follow our full recipe with a single unit at the higher levels, and with setting the higher level design parameters for variable importance and variation to 0.

3.1 Determine DGP parameters

We have already discussed most of the required parameters in Section 1. However, there are a few additional parameters required to generate data that do not directly feed into our equations for single-outcome power or MDES. These are the correlations shown in Table 7.

The parameters used in this section need to be picked based on desired aggregate relationships of the full data. See the next section for how to translate parameters such as ICC to the DGP parameters

Param	Description
$\boldsymbol{\rho}^V$	Correlation matrix of district covariates $\boldsymbol{V}_{k.}$
$\boldsymbol{\rho}^{w_0}$	Correlation matrix of district random effects $\boldsymbol{w}_{0,k.}$
$\boldsymbol{\rho}^{w_1}$	Correlation matrix of district impacts $\boldsymbol{w}_{1,k.}$
$\boldsymbol{\kappa}^w$	Diagonal matrix of correlations between district random effects and impacts, composed of entries $\{\kappa_m^w\} = Corr(w_{0,km}, w_{1,km})$
$\boldsymbol{\rho}^X$	Correlation matrix of school covariates $\boldsymbol{X}_{jk.}$
$\boldsymbol{\rho}^{u_0}$	Correlation matrix of school random effects $\boldsymbol{u}_{0,jk.}$
$\boldsymbol{\rho}^{u_1}$	Correlation matrix of school impacts $\boldsymbol{u}_{1,jk.}$
$\boldsymbol{\kappa}^u$	Diagonal matrix of correlations between school random effects and impacts, composed of entries $\{\kappa_m^u\} = Corr(u_{0,jkm}, u_{1,jkm})$
$\boldsymbol{\rho}^C$	Correlation matrix of individual covariates $\boldsymbol{C}_{ijk.}$
$\boldsymbol{\rho}^r$	Correlation matrix of individual residuals $\boldsymbol{r}_{ijk.}$

Table 7: Correlation parameters.

3.2 Generate level 3 (district) data

3.2.1 Level 3 covariates

Each outcome has its own district-level covariate, V_{km} with $k = 1, \ldots, K$ and $m = 1, \ldots, M$. For example, district average reading and math pre-tests, used for adjusting reading and math outcomes, respectively (in practice we might imagine adjusting each outcome with both, but in the case of few districts this might not be a good idea due to degrees of freedom issues). We have $E(V_{km}) = 0$ and $Var(V_{km}) = 1$. We assume a correlation between the covariates across outcomes, so we define $\boldsymbol{\rho}^V$ as a $M \times M$ symmetric correlation matrix, with ρ^V_{ij} the value in row i and column j.

$$\begin{pmatrix} V_{k1} \\ \vdots \\ V_{km} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \cdots & \rho^V_{1M} \\ \vdots & 1 & \vdots \\ \rho^V_{M1} & \cdots & 1 \end{pmatrix} \right]. \tag{102}$$

3.2.2 Level 3 outcomes

Let $\psi_{0,km}$ be the grand mean outcome under no treatment for district k, and $\psi_{1,km}$ be the grand mean impact across schools for district k. We define them in terms of overall grand means with district-specific offsets:

$$\psi_{0,km} = \Xi_{0,m} + \xi_m V_{km} + w_{0,km} \tag{103}$$

$$\psi_{1,km} = \Xi_{1,m} + w_{1,km}, \tag{104}$$

where $\Xi_{0,m}$ is the grand mean outcome under no treatment across all districts, and $\Xi_{1,km}$ is the grand mean impact across districts. Without loss of generality, we will set $\Xi_{0,m} = 0$ for all m. $\Xi_{1,km}$ has to be set by the user.

We generate the random effects $w_{0,km}$ and $w_{1,km}$ via a multivariate normal distribution, and then calculate the $\psi_{0,km}$ and $\psi_{1,km}$ by plugging everything in to the above equations.

To generate the random effects, we first specify $E(w_{0,km}) = 0$ and $Var(w_{0,km}) = \eta^2_{0,m}$. We then set the pairwise correlation between $w_{0,km}$ and $w_{0,km'}$ for outcomes m and m' as $\rho^{w_0}_{mm'}$. We store the full set of correlations in a correlation matrix $\boldsymbol{\rho}^{w_0}$ (with a 1 diagonal).

This gives our m-vectors of random district-level intercepts as multivariate normal:

$$\begin{pmatrix} w_{0,k1} \\ \vdots \\ w_{0,kM} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2_{0,1} & \cdots & \rho^{w_0}_{1M}\eta_{0,1}\eta_{0,M} \\ \vdots & \ddots & \vdots \\ \rho^{w_0}_{M1}\eta_{0,M}\eta_{0,1} & \cdots & \eta^2_{0,M} \end{pmatrix} \right] = N(\vec{0}, \Sigma_{w_0}). \tag{105}$$

Similarly, we have $E(w_{1,km}) = 0$, $Var(w_{1,km}) = \eta^2_{1,m}$, and an $M \times M$ correlation matrix $\boldsymbol{\rho}^{w_1}$, giving a distribution for the random impact residuals of $N(\vec{0}, \Sigma_{w_1})$:

$$\begin{pmatrix} w_{1,k1} \\ \vdots \\ w_{1,kM} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2_{1,1} & \cdots & \rho^{w_1}_{1M}\eta_{1,1}\eta_{1,M} \\ \vdots & \ddots & \vdots \\ \rho^{w_1}_{M1}\eta_{1,M}\eta_{1,1} & \cdots & \eta^2_{1,M} \end{pmatrix} \right] = N(\vec{0}, \Sigma_{w_1}). \tag{106}$$

We also need to specify the correlation of the random intercepts with the random impacts within each outcome. In particular, for outcome m, we want the joint distribution, $(w_{0,km}, w_{1,km})$ to be bivariate normal with correlation κ_m^w:

$$\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta_{0,m}^2 & \kappa_m^w \eta_{0,m}\eta_{1,m} \\ \kappa_m^w \eta_{1,m}\eta_{0,m} & \eta_{1,m}^2 \end{pmatrix} \right], \tag{107}$$

We can also collect the κ_m^w into a diagonal matrix $\boldsymbol{\kappa}^w$.

The three distributions above can be combined into a large $2M$-dimensional multivariate gaussian distribution:

$$(w_{0,k1}, \ldots, w_{0,kM}, w_{1,k1}, \ldots, w_{1,km}) \sim MVNorm(\vec{0}, \Sigma_{full}^w), \tag{108}$$

with

$$\Sigma_{full}^w = \begin{pmatrix} \Sigma_{w_0} & \Sigma_w \\ \Sigma_w' & \Sigma_{w_1} \end{pmatrix}. \tag{109}$$

The Σ_w contains the intercept-impact covariances in Equation 107, but also have other elements. In Section 5.2 we derive these, assuming a working linear model of the individual elements. This gives Σ_w as

$$\Sigma^w = \begin{pmatrix} \kappa_1^w \eta_{0,1}\eta_{1,1} & \cdots & \frac{1}{2}\left(\kappa_1^w \rho_{1M}^{w_1} + \kappa_M^w \rho_{1M}^{w_0}\right)\eta_{0,1}\eta_{1,M} \\ \vdots & \ddots & \vdots \\ \frac{1}{2}\left(\kappa_M^w \rho_{M1}^{w_1} + \kappa_1^w \rho_{M1}^{w_0}\right)\eta_{0,M}\eta_{1,1} & \cdots & \kappa_M^w \eta_{0,M}\eta_{1,M} \end{pmatrix}. \tag{110}$$

Note how the diagonals correspond to the off-diagonal in Eq 107; this is what gives us the proper pairwise correlation of random intercept and random impact residual within each outcome. In most situations, we set $\boldsymbol{\kappa}^w = \mathbf{0}$, so Σ^w is a matrix of all 0.

3.3 Generate level 2 (school) data

The schools' X_{jkm}, $\theta_{0,jkm}$, $\theta_{1,jkm}$ are all generated in the same way the district outcomes are generated. Just as with districts, each outcome has its own school-level covariate.

As with the district-level covariates, we have $E(X_{jkm}) = 0$ and $Var(X_{jkm}) = 1$, and $\boldsymbol{\rho}^X$ is a $M \times M$ symmetric correlation matrix that allows generation of the covariates.

Each school j in district k has its average outcome under no treatment $\theta_{0,jkm}$ and its average impacts $\theta_{1,jkm}$, and they are generated by first generating all the random effects, and then plugging everything into

$$\theta_{0,jkm} = \psi_{0,km} + \delta_m X_{jkm} + u_{0,jkm} \tag{111}$$
$$\theta_{1,jkm} = \psi_{1,km} + u_{1,jkm}. \tag{112}$$

The $(u_{0,jkm}, u_{1,jkm})$ follow a multivariate Normal structure as in Section 3.2.2. We have $Var(u_{0,jkm}) = \tau_{0,m}^2$ and $Var(u_{1,jkm}) = \tau_{1,m}^2$. Also $Cov(\boldsymbol{u}_{0,jk\cdot}) = \boldsymbol{\rho}^{u_0}$ and $Cov(\boldsymbol{u}_{1,jk\cdot}) = \boldsymbol{\rho}^{u_1}$. Finally, they relate to each other with $Corr(u_{0,jkm}, u_{1,jkm}) = \kappa_m^u$. Thus, we have a matrix Σ^u that is constructed like Σ^w, where $\Sigma_{m,m'}^u = \kappa_m^u \rho_{mm'}^{u_1} \tau_{0,m} \tau_{1,m'}$.

We can easily convert from three-level to two-level models. If there are no districts, then $\psi_{0,km} = \Xi_{0,m}$ and $\psi_{1,km} = \Xi_{1,m}$ for all k. Essentially, we set $w_{km} = 0$, $w_{km} = 0$, and $\xi_m = 0$ for all k.

3.4   Generate level 1 (individual) data

3.4.1   Level 1 covariates

Individuals have individual level covariates, one for each outcome C_{ijkm}. For example, group-mean centered reading and math scores. We assume these are homoskedastic and have the same mean across sites. As with previous covariates, we have $E(C_{ijkm}) = 0$ and $Var(C_{ijkm}) = 1$, and $\boldsymbol{\rho}^C$ is a $M \times M$ symmetric correlation matrix.

3.4.2   Level 1 outcomes

For each outcome, the outcome model for the individual is

$$Y_{ijkm}(0) = \theta_{0,jkm} + \gamma_m C_{ijkm} + r_{ijkm} \tag{113}$$
$$Y_{ijkm}(1) = Y_{ijkm}(0) + \theta_{1,ijkm}, \tag{114}$$

where $Y_{ijkm}(0)$ is potential outcome m under no treatment for individual i in school j in district k, and θ_{ijkm} is the unit's individual causal effect.

We assume constant treatment effects for individuals in the same school, $\theta_{ijkm} = \theta_{1,jkm}$, but this assumption could be relaxed to allow for individual treatment-level heterogeneity.

As with previous covariates, we have $E(C_{ijkm}) = 0$ and $Var(C_{ijkm}) = 1$, and $\boldsymbol{\rho}^C$ is a $M \times M$ symmetric correlation matrix.

Finally, individual-level residuals are distributed $E(r_{ijkm}) = 0$ and $Var(r_{ijkm}) = 1$, and $\boldsymbol{\rho}^r$ is a $M \times M$ symmetric correlation matrix.

3.4.3   Reduced form

Putting the levels together, we have:

$$Y_{ijkm}(0) = \Xi_{0,m} + \xi_m V_{km} + \delta_m X_{jkm} + \gamma_m C_{ijkm} + w_{0,km} + u_{0,jkm} + r_{ijkm} \tag{115}$$
$$Y_{ijkm}(1) = Y_{ijkm}(0) + \Xi_{1,m} + w_{1,km} + u_{1,jkm}. \tag{116}$$

3.5   Summary: Generating the full table of potential outcomes

1. For $k = 1, \ldots K$, and $m = 1, \ldots M$:

   (a) Generate district covariates V_{km}.

   (b) Generate district residuals $w_{0,km}$ and $w_{1,km}$.

   (c) Calculate district grand means $\psi_{0,km}$ and average impacts $\psi_{1,km}$.

2. For $j = 1, \ldots J$, and $m = 1, \ldots M$:

   (a) Generate school covariates X_{jkm}.

   (b) Generate school residuals $u_{0,jkm}$ and $u_{1,jkm}$.

   (c) Calculate school grand means $\theta_{0,jkm}$ and average impacts $\theta_{1,jkm}$.

3. For $i = 1, \ldots N$ and for $m = 1, \ldots M$:

(a) Generate individual covariates, C_{ijkm}

(b) Generate individual residuals r_{ijkm}.

(c) Generate predicted baseline outcomes ($Y_{ijkm}(0)$ without residuals).

(d) Add residuals to the predicted outcomes to get $Y_{ijkm}(0)$

(e) Calculate $Y_{ijkm}(1)$ by adding the treatment model to the $Y_{ijkm}(0)$.

3.6 Generate observed data

Once we have our full set of potential outcomes, we generate treatment assignments to generate the observed outcomes. We generate our treatment assignment, T_{ijk} for all $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Once we have our set of T_{ijk} (no matter how they were obtained) we calculate the observed outcomes

$$Y_{ijkm}^{obs} = Y_{ijkm}(0)(1 - T_{ijk}) + Y_{ijkm}(1)T_{ijk}. \tag{117}$$

3.6.1 Randomization schemes

We can assign at the district, school, or individual level depending on the design we are generating data for.

- Blocked individual randomization: simple random sampling occurs within each school, with $\bar{n}\bar{T}$ units assigned to treatment in each school.

- Cluster 2-level randomization: simple random sampling occurs across schools, with $J\bar{T}$ schools assigned to treatment.

- Blocked cluster 2-level randomization: school level assignment occurs within each district, with $J\bar{T}$ schools assigned to treatment in each district.

- Cluster 3-level randomization: simple random sampling occurs across districts, with $K\bar{T}$ districts assigned to treatment.

4 Tuning the DGP parameters

We define two main classes of parameters: model parameters and control parameters. Model parameters, listed in Tables 1, 2, 3 and 7, are the ones that define the DGP. We use them directly to simulate data. The control parameters, defined in Table 6, are more interpretable and can be used to indirectly tune (i.e. specify) the model parameters. In short, control parameters are set by the user, and then are converted to the model parameters that are fed into the DGP.

We break our model parameters into sets:

- Set 1: $\{M, J, K, n_{jk}, \Xi_{0,m}, \boldsymbol{\rho}^D, \boldsymbol{\rho}^w, \boldsymbol{\rho}^z, \boldsymbol{\rho}^X, \boldsymbol{\rho}^u, \boldsymbol{\rho}^v, \boldsymbol{\rho}^C, \boldsymbol{\kappa}^{wz}, \boldsymbol{\kappa}^{uv}, p_j\}$ are set directly by the user.

- Set 2: $\{N, \psi_{0,km}, \psi_{1,km}, \theta_{0,jkm}, \theta_{1,jkm}, Y_{ijkm}(0), Y_{ijkm}(1)\}$ are functions of parameters that are set directly.

- Set 3: $\{\Xi_{1,m}, \eta_{0,m}^2, \eta_{1,m}^2, \tau_{0,m}^2, \tau_{1,m}^2, \xi_m, \delta_m, \gamma_m\}$ are tuned through the control parameters.

The mechanism that maps the control parameters introduced in Section 1.5 to the model parameters come from several derived relationships that we provide in the following sections.

4.1 Calculating the variation in random effects and impacts

We can have variation at the individual, school, and district level. We want to be able to tune the proportion of variation in each of these levels. We are interested in the expression for unconditional (covariate-free) ICC on the control side.

Using Equation 115, we have the variance of the control side as:

$$Var_m(Y_{ijkm}(0)) = \xi_m^2 Var_m(V_{km}) + \delta_m^2 Var_m(X_{jkm}) + \gamma_m^2 Var_m(C_{ijkm}) + \eta_{0,m}^2 + \tau_{0,m}^2 + \sigma_m^2$$
$$= \xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + \sigma_m^2.$$

assuming all covariates have a variance of 1.

Which leads to our definitions of ICC, as previously defined in Equations 11-12.

$$\text{ICC}_{3,m} = \frac{Var(\psi_{0,km})}{Var(Y_{ijkm}(0))} = \frac{\xi_m^2 + \eta_{0,m}^2}{\left(\xi_m^2 + \eta_{0,m}^2\right) + \left(\delta_m^2 + \tau_{0,m}^2\right) + \left(\gamma_m^2 + \sigma_m^2\right)}$$

$$\text{ICC}_{2,m} = \frac{Var(\theta_{0,jkm})}{Var(Y_{ijkm}(0))} = \frac{\delta_m^2 + \tau_{0,m}^2}{\left(\xi_m^2 + \eta_{0,m}^2\right) + \left(\delta_m^2 + \tau_{0,m}^2\right) + \left(\gamma_m^2 + \sigma_m^2\right)}.$$

4.2 Calculating the covariate coefficients

4.2.1 Calculating the level 3 covariate coefficient ξ_m

The regression coefficients for the level 3 covariates, ξ_m, are dictated by the desired R^2 values. Thus, we would like to find ξ_m as a function of $R_{3,m}^2$, i.e., the proportion of variance between level 3 districts predicted by level 3 covariates.

We start with

$$R_{3,m}^2 = 1 - \frac{Var(w_{0,km})}{Var(\psi_{0,km})}$$

$$= 1 - \frac{\eta_{0,m}^2}{\xi_m^2 Var(V_{km}) + \eta_{0,m}^2},$$

leading to

$$\xi_m = \sqrt{\frac{\eta_{0,m}^2 R_{3,m}^2}{Var(V_{km})(1 - R_{3,m}^2)}}$$

$$= \sqrt{\frac{\eta_{0,m}^2 R_{3,m}^2}{1 - R_{3,m}^2}}.$$

4.2.2 Calculating the level 2 covariate coefficient δ_m

We start with our level-2 R^2 being defined as the proportion of variance in level-2 schools explained by level-2 covariates:

$$R_{2,m}^2 = 1 - \frac{Var(u_{0,jkm})}{Var(\theta_{0,jkm} \mid D_{id})},$$

where the conditioning $Var(\theta_{0,jkm} \mid D_{id})$ denotes the variance of outcomes *within* a particular district. This expands to

$$R_{2,m}^2 = 1 - \frac{\tau_{0,m}^2}{\delta_m^2 Var(X_{jkm} \mid D_{id}) + \tau_{0,m}^2}.$$

Since our X_{jkm} are generated independent of district, the conditional variance is the same as the overall variance. This gives

$$R_{2,m}^2 = 1 - \frac{\tau_{0,m}^2}{\delta_m^2 Var(X_{jkm}) + \tau_{0,m}^2},$$

leading to

$$\delta_m = \sqrt{\frac{\tau_{0,m}^2 R_{1,m}^2}{Var(X_{jkm})(1 - R_{2,m}^2)}}$$

$$= \sqrt{\frac{\tau_{0,m}^2 R_{2,m}^2}{1 - R_{2,m}^2}}.$$

4.2.3   Calculating the coefficient for the Level 1 variable (γ_m)

Similar to level 2, we start with our level 1 R^2 being defined as the proportion of level 1 variance in individuals explained by level 1 covariates:

$$R_{1,m}^2 = 1 - \frac{\sigma_m^2}{var(Y_{ijkm}(0) \mid S_{id})},$$

where the conditioning denotes the variance of outcomes *within* a particular school.

We find

$$R_{1,m}^2 = 1 - \frac{\sigma_m^2}{\gamma_m^2 var(C_{ijkm} \mid S_{id}) + \sigma_m^2}$$

$$= 1 - \frac{\sigma_m^2}{\gamma_m^2 var(C_{ijkm}) + \sigma_m^2}.$$

Rearrange to get

$$\gamma_m = \sqrt{\frac{\sigma_m^2 R_{1,m}^2}{var(C_{ijkm})(1 - R_{1,m}^2)}}$$

$$= \sqrt{\frac{R_{1,m}^2}{1 - R_{1,m}^2}}.$$

4.3 Calculating the grand mean impacts $\Xi_{1,m}$

The grand mean impact is a function of effect size. The effect size is simply the overall impact in standard deviation units, with the standard deviation usually being the marginal standard deviation of the control side:

$$\Xi_{1,m} = ES_m \cdot SD_m(Y_{ijkm}(0)), \tag{118}$$

where $SD_m(Y_{ijkm}(0))$ denotes the standard deviation over i, j, and k for fixed outcome m. We have already noted $Var_m(Y_{ijkm}(0)) = \xi_m^2 + \gamma_m^2 + \delta_m^2 + \eta_{0,m}^2 + \tau_{0,m}^2 + \sigma_m^2$.

4.4 Final results

The above produce the following system of equations:

$$\text{ICC}_{3,m} = \frac{\xi_m^2 + \eta_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + 1}$$

$$\text{ICC}_{2,m} = \frac{\delta_m^2 + \tau_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + 1}$$

$$\xi_m = \sqrt{\frac{\eta_{0,m}^2 R_{3,m}^2}{1 - R_{3,m}^2}}$$

$$\delta_m = \sqrt{\frac{\tau_{0,m}^2 R_{2,m}^2}{1 - R_{2,m}^2}}$$

$$\gamma_m = \sqrt{\frac{R_{1,m}^2}{1 - R_{1,m}^2}}.$$

We solve the system to find our model parameters in terms of our user parameters:

$$\gamma_m^2 = \frac{R_{1,m}^2}{1 - R_{1,m}^2} \tag{119}$$

$$\delta_m^2 = \frac{R_{2,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{2,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}} \tag{120}$$

$$\xi_m^2 = \frac{R_{3,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{3,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}} \tag{121}$$

$$\tau_{0,m}^2 = \frac{1 - R_{2,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{2,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}} \tag{122}$$

$$\eta_{0,m}^2 = \frac{1 - R_{3,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{3,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}}. \tag{123}$$

For details on the algebra, see Section 5.

And finally we set:

$$\eta_{1,m}^2 = \omega_{3,m} \left(\eta_{0,m}^2 + \xi_m^2 \right) \tag{124}$$

$$\tau_{1,m}^2 = \omega_{2,m} \left(\tau_{0,m}^2 + \delta_m^2 \right). \tag{125}$$

5 Derivations

5.1 Derivations of parameter translation formulae

In this section we derive the formulae used to map user parameters to model parameters given above. We first lay out some expressions we will later use in the main derivations:

$$\tau_{0,m}^2 = \frac{\delta_m^2(1 - R_{2,m}^2)}{R_{2,m}^2}$$

$$\delta_m^2 + \tau_{0,m}^2 = \delta_m^2 + \frac{\delta_m^2(1 - R_{2,m}^2)}{R_{2,m}^2}$$

$$= \frac{\delta_m^2 R_{2,m}^2 + \delta_m^2 - \delta_m^2 R_{2,m}^2}{R_{2,m}^2},$$

giving

$$\delta_m^2 + \tau_{0,m}^2 = \frac{\delta_m^2}{R_{2,m}^2}.$$

We also note:

$$\frac{\text{ICC}_{3,m}}{\text{ICC}_{2,m}} = \frac{\xi_m^2 + \eta_{0,m}^2}{\delta_m^2 + \tau_{0,m}^2}$$

$$\xi_m^2 + \eta_{0,m}^2 = \frac{\text{ICC}_{3,m}(\delta_m^2 + \tau_{0,m}^2)}{\text{ICC}_{2,m}}$$

$$= \frac{\text{ICC}_{3,m}\delta_m^2}{R_{2,m}^2 \text{ICC}_{2,m}}.$$

And finally it is easy to re-express $\gamma_m^2 + 1$:

$$\gamma_m = \sqrt{\frac{R_{1,m}^2}{1 - R_{1,m}^2}}$$

$$\gamma_m^2 + 1 = \frac{R_{1,m}^2}{1 - R_{1,m}^2} + 1$$

$$= \frac{1}{1 - R_{1,m}^2}.$$

Let's start by plugging some of these into our expression for ICC_2 to find δ_m:

$$\mathrm{ICC}_{2,m} = \frac{\delta_m^2 + \tau_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + 1}$$

$$= \frac{\dfrac{\delta_m^2}{R_{2,m}^2}}{\dfrac{\mathrm{ICC}_{3,m}\delta_m^2}{R_{2,m}^2 \mathrm{ICC}_{2,m}} + \dfrac{\delta_m^2}{R_{2,m}^2} + \gamma_m^2 + 1}$$

$$\frac{\delta_m^2}{R_{2,m}^2} = \mathrm{ICC}_{2,m}\left(\frac{\mathrm{ICC}_{3,m}\delta_m^2}{R_{2,m}^2 \mathrm{ICC}_{2,m}} + \frac{\delta_m^2}{R_{2,m}^2} + \gamma_m^2 + 1\right)$$

$$\delta_m^2 = \mathrm{ICC}_{3,m}\delta_m^2 + \mathrm{ICC}_{2,m}\delta_m^2 + \mathrm{ICC}_{2,m}R_{2,m}^2(\gamma_m^2 + 1)$$

$$\delta_m^2 = \frac{\mathrm{ICC}_{2,m}R_{2,m}^2(\gamma_m^2 + 1)}{1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m}}$$

$$= \frac{\mathrm{ICC}_{2,m}R_{2,m}^2}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}.$$

Proceeding by a similar method, we can use ICC_3 to find ξ_m:

$$\xi_m^2 = \frac{\mathrm{ICC}_{3,m}R_{3,m}^2}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}.$$

Now we can plug in to find $\tau_{0,m}^2$:

$$\tau_{0,m}^2 = \frac{\delta_m^2(1 - R_{2,m}^2)}{R_{2,m}^2}$$

$$= \frac{\mathrm{ICC}_{2,m}R_{2,m}^2}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)} \frac{(1 - R_{2,m}^2)}{R_{2,m}^2}$$

$$= \frac{\mathrm{ICC}_{2,m}(1 - R_{2,m}^2)}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}.$$

And similarly:

$$\eta_{0,m}^2 = \frac{\mathrm{ICC}_{3,m}(1 - R_{3,m}^2)}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}.$$

5.2 Derivation of off-diagonal correlations of outcomes

We directly specify most of the $2M \times 2M$ covariance matrix of the random effects for districts and schools. The off-diagonal terms, however, need to be derived as natural consequences of these choices. One could specify these off-diagonal terms directly, as long as the resulting matrix were positive definite, but we can also view them as artifacts of the other correlations coupled with a simplifying linear model. That is what we do here.

Remember that we define κ_m^w as the correlation of the intercept and impact for outcome m, which corresponds to the correlation for the diagonals. We can also collect the κ_m^w into a diagonal matrix $\boldsymbol{\kappa}^w$.

In particular, consider the correlation $cor(w_{0,km}, w_{1,km'})$ between the random effect for outcome m and random impact residual for outcome m'. If there is a non-zero correlation between $(w_{0,km}, w_{1,km})$, and a non-zero correlation between $(w_{1,km}, w_{1,km'})$, then, unless something is offsetting it, we will have a non-zero correlation between $(w_{0,km}, w_{1,km'})$.

In path analysis, we consider all paths that could induce correlations between two variables $(w_{0,km}, w_{1,km'})$. One path is $w_{0,km}$—$w_{1,km}$—$w_{1,km'}$. Another path is $w_{0,km}$—$w_{0,km'}$—$w_{1,km'}$. Below, we show the derivation for the correlation due to the first path. We do not show the derivation for the second path, as it follows the same pattern.

Let's assume there is a linear relationship between the two pairs of variables with known correlations.

$$w_{0,km} = a + bw_{1,km} + C$$
$$w_{1,km} = d + ew_{1,km'} + F,$$

where C and F are error terms.

First, let's find the definitions of b and e in terms of notation we have already defined: $\rho_{mm'}^{w_1}$, κ_m^w, $\eta_{0,m}^2$, and $\eta_{1,m}^2$.

5.2.1   Finding the linear coefficients.

Finding b.   First, we find b in terms of our original notation. Based on the linear relationships, we have:

$$
\begin{aligned}
Cov(w_{0,km}, w_{1,km}) &= Cov(a + bw_{1,km} + C, w_{1,km}) \\
&= bCov(w_{1,km}, w_{1,km}) + Cov(C, w_{1,km}) \\
&= b\eta_{1,m}^2.
\end{aligned}
$$

Now to relate b to our original notation, we defined $Cor(w_{0,km}, w_{1,km}) = \kappa_m^w$, and $Cov(w_{0,km}, w_{1,km}) = \kappa_m^w \eta_{0,m} \eta_{1,m}$. This means we can solve for b in terms of our original variables:

$$b\eta_{1,m}^2 = \kappa_m^w \eta_{0,m} \eta_{1,m}$$
$$b = \frac{\kappa_m^w \eta_{0,m}}{\eta_{1,m}}.$$

Finding e.   Next, we find e. Based on the linear relationships, we have:

$$
\begin{aligned}
Cov(w_{1,km}, w_{1,km'}) &= Cov(d + ew_{1,km'} + F, w_{1,km'}) \\
&= eCov(w_{1,km'}, w_{1,km'}) + Cov(F, w_{1,km'}) \\
&= e\eta_{1,m'}^2.
\end{aligned}
$$

Now to relate e to our original notation, we defined $Cor(w_{1,km}, w_{1,km'}) = \rho_{mm'}^{w_1}$, and $Cov(w_{1,km}, w_{1,km'}) = \rho_{mm'}^{w_1} \eta_{1,m} \eta_{1,m'}$. This means we can solve for e in terms of our original variables:

$$e\eta_{1,m'}^2 = \rho_{mm'}^{w_1} \eta_{1,m} \eta_{1,m'}$$
$$e = \frac{\rho_{mm'}^{w_1} \eta_{1,m}}{\eta_{1,m'}}.$$

5.2.2 Finding the covariance expression.

Next, we re-arrange our expressions to be in terms of the common variable $w_{1,km'}$.

$$w_{0,km} = a + b(d + ew_{1,km'} + F) + C$$

Now, we plug in our new expression to find the covariance.

$$
\begin{aligned}
Cov(w_{0,km}, w_{1,km'}) &= Cov\left(a + bd + bew_{1,km'} + bF + C, w_{1,km'}\right) \\
&= Cov\left(bew_{1,km'} + bF + C, w_{1,km'}\right) \\
&= beCov(w_{1,km'}, w_{1,km'}) + bCov(F, w_{1,km'}) + Cov(C, w_{1,km'}) \\
&= beCov(w_{1,km'}, w_{1,km'}) \\
&= be\eta_{1,m'}^2.
\end{aligned}
$$

Finally, we substitute in our values of b and e.

$$
\begin{aligned}
Cov(w_{0,km}, w_{1,km'}) &= \frac{\kappa_m^w \eta_{0,m}}{\eta_{1,m}} \frac{\rho_{mm'}^{w_1} \eta_{1,m}}{\eta_{1,m'}} \eta_{1,m'}^2 \\
&= \kappa_m^w \rho_{mm'}^{w_1} \eta_{0,m} \eta_{1,m'}.
\end{aligned}
$$

5.2.3 Second path.

As mentioned above, another path is $w_{0,km}$—$w_{0,km'}$—$w_{1,km'}$. Considering this path, by similar derivation to the first path, we would find:

$$Cov(w_{0,km}, w_{1,km'}) = \kappa_{m'}^w \rho_{mm'}^{w_0} \eta_{0,m} \eta_{1,m'}.$$

5.2.4 Total covariance.

To find the total covariance, we take the average covariance over the two paths. We take the average so that the diagonals have the correct correlation κ_m^w, which we already defined as the correlation of the intercept and impact for outcome m.

$$Cov(w_{0,km}, w_{1,km'}) = \frac{1}{2}\left(\kappa_m^w \rho_{mm'}^{w_1} + \kappa_{m'}^w \rho_{mm'}^{w_0}\right) \eta_{0,m} \eta_{1,m'}. \tag{126}$$

5.2.5 Constructing the non-symmetric matrix Σ^w.

We can now construct our full matrix Σ^w, noting that for $m = m'$, $\rho_{mm'}^{w_1} = \rho_{mm'}^{w_0} = 1$:

$$\Sigma^w = \begin{pmatrix} \kappa_1^w \eta_{0,1} \eta_{1,1} & \cdots & \frac{1}{2}\left(\kappa_1^w \rho_{1M}^{w_1} + \kappa_M^w \rho_{1M}^{w_0}\right) \eta_{0,1} \eta_{1,M} \\ \vdots & \ddots & \vdots \\ \frac{1}{2}\left(\kappa_M^w \rho_{M1}^{w_1} + \kappa_1^w \rho_{M1}^{w_0}\right) \eta_{0,M} \eta_{1,1} & \cdots & \kappa_M^w \eta_{0,M} \eta_{1,M} \end{pmatrix}. \tag{127}$$

References

Howard S. Bloom. The core analytics of randomized experiments for social research, 2006.

M. Bulus, N. Dong, B. Kelcey, and J Spybrook. PowerUpR: Power analysis tools for multilevel randomized experiments. r package version 1.1.0. [software]., 2022. URL https://CRAN.R-project.org/package=PowerUpR.

Nianbo Dong and Rebecca Maynard. PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6(1):24–67, 2013. ISSN 1934-5747.

Luke W Miratrix, Michael J Weiss, and Brit Henderson. An applied researcher's guide to estimating effects from multisite individually randomized trials: Estimands, estimators, and estimates. *Journal of Research on Educational Effectiveness*, 14(1):270–308, 2021.

J Spybrook, Larry Hedges, and Michael Borenstein. Understanding statistical power in cluster randomized trials: Challenges posed by differences in notation and terminology. *Journal of Research on Educational Effectiveness*, 7:384–406, 2014a.

Jessaca Spybrook, Larry Hedges, and Michael Borenstein. Understanding statistical power in cluster randomized trials: Challenges posed by differences in notation and terminology. *Journal of Research on Educational Effectiveness*, 7(4):384–406, 2014b.